Mukesh Patel School of Technology Management and Engineering
Computer Engineering Department

Course Policy

Program/Branch/Semester	**:**	B.Tech/MBATECH (Computer,EXTC)
Academic Year	**:**	2023-24
Course Code & Name	**:**	Machine Learning
Credit Details	**:**	L T P C 2 0 2 3
Course Coordinator Faculty	**:**	
Contact No. & Email	**:**	
Office	**:**	
Contact hours	**:**	Saturday (10am-12pm
Other Course Faculty members teaching this course	**:**	Course Faculty 1: Dr .Shubha Puthran Contact No. & Email: Office: Office Hours: Saturday (10am-12pm) Course Faculty 2: Dr.Prashasti Kanikar Contact No. & Email: Office: Office Hours: Saturday (10am-2pm) Course Faculty 3: Prof.Abhay Kohle Contact No. & Email: Office: Office Hours: Saturday (10am-12pm) Course Faculty 4: Dr.Sandip Contact No. & Email: Office:
Queries by Emails are encouraged.		
Course link	**:**	Portal Link:- https://portal.svkm.ac.in/MPSTME-NM-M/homepage MS Teams Link :- As submission is to be conducted on Student Portal, it is not applicable.

1 Introduction to the Course

1.1 Importance of the course

Machine Learning is a contemporary, practical oriented, application based course. It is a subset of artificial intelligence where systems can learn from data, learn patterns and make decisions. Supervised machine learning algorithms like regression and classification and unsupervised algorithms like clustering can be applied in multiple areas. Business analytics, banking sectors, medical fields, industry automation are to name the few.

1.2 Objective of the Course

The principal objective of this course is a concise introduction to the fundamental concepts in machine learning and machine learning algorithms from a practical perspective which will help students to solve real world problems.

1.3 Pre-requisite

- Knowledge of Probability and statistics, Calculus and Linear Algebra.
- Python Programming

2 Course Outcomes (CO) and mapping with Program Outcomes (PO)

2.1 Course Outcomes

After successful completion of the course, a student will be able to-

1. Identify machine learning techniques suitable for a given problem.
2. Apply regression model to solve problems in machine learning domain.
3. Analyze and apply classification and clustering algorithms on data

2.2 CO-PO Mapping

	PO 1	PO2	PO3	PO4	PO 5	PO 6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2
CO1		3	3		3		2		M			
CO2	3		2	2		3		3			3	
CO3			3	3	3						3	M

3-High Mapping 2- Medium mapping 1-Low mapping

3 Syllabus, Pre-class activity and References

3.1 Teaching and evaluation scheme

Teaching Scheme				Evaluation Scheme	
Lecture Hours per week	**Practical Hours per week**	**Tutorial Hours per week**	**Credit**	**Internal Continuous Assessment (ICA) As per Institute Norms (50 Marks)**	**Theory (3 Hrs, 100 Marks)**
2	2	0	3	Marks Scaled to 50	Marks Scaled to 50

3.2 Syllabus

Detailed Syllabus:		
Unit	**Description**	**Duration**
1.	**Machine Learning Fundamentals** Terminology, Supervised and Unsupervised Learning with examples, Underfitting / Overfitting, Bias-Variance Trade-off, Model Selection, Applications	02
2.	**Exploratory data Analysis** Missing Value Treatment, Handling Categorical data: Mapping ordinal features, Encoding class labels, Performing one-hot encoding on nominal features, Outlier Detection and Treatment. Feature Engineering: Variable Transformation and Variable Creation, Selecting meaningful features.	02
3.	**Regression** Linear regression using Least Squares (analytical approach), Linear regression using Gradient Descent (iterative approach), Multiple linear regression, Polynomial regression	06
4.	**Classification** Performance Evaluation, Confusion Matrix, Accuracy, Precision, Recall, F1-score, ROC Curves, AUC, k-fold Cross-Validation. Logistic Regression, Naive Bayes Classifier, Support Vector Machines, Neural Networks: Perceptron, Multi-layer Perceptron, Training using Back-propagation, Applications	08
5.	**Tree-Based Methods** Basics of Decision Trees, Regression Trees, Classification Trees, Trees v/s Linear Models, Advantages and Disadvantages of Trees, Ensemble techniques- Bagging, Boosting, Random forest	06
6.	**Unsupervised Learning** Challenges of Unsupervised Learning, Partitioning Methods: K-Means clustering, Dimension Reduction Methods, Principal	06

	Component Analysis (PCA), Hierarchical Clustering, Introduction to Recommender systems (Non-personalized and Content-based)	
	Total	**30**

Text Books:

1. Judith Hurwitz, Daniel Kirsch, "Machine Learning for dummies", IBM Limited Edition, John Wiley & Sons, Inc., 2018.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", 2nd Edition, Springer, 2017.

Reference books

1. Aurelien Geron, "Hands-on Machine Learning with Scikit Learn, Keras and Tensorflow", 2nd Edition, Oreilly Publication, 2019.
2. James, Witten, Hastie, Tibshirani," Introduction to Statistical Learning", 7th Edition, Springer, 2017.

3.3 Pre-class activity

Outline for preliminary study to be done for each unit will be provided prior to commencement of each unit. Preliminary study material (video links, presentation, notes etc.) will be made available on the student portal. Students are expected to go through this material before attending the upcoming session. It is expected that the students put in at least two hours of self-study for every one hour of classroom teaching. During the lecture session, more emphasis will be given on in-depth topics, practical applications and doubt solving.

3.4 References

Text Books:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of StatisticalLearning*: *Data Mining, Inference, and Prediction,* Second Edition, Springer, 2017.
2. Aurelien Geron, *Hands-on Machine Learning with Scikit Learn, Keras and Tensorflow* ,2nd Edition, Oreilly Publication, 2019.

Reference books:

1. Kavin Murphy, *Machine Learning: A probabilistic perspective*, MIT press, 2012.
2. Judith Hurwitz, Daniel Kirsch, *Machine Learning for dummies*, IBM Limited Edition,John Wiley & Sons, Inc., 2018.

Note: The latest edition of books should be referred.

4 Laboratory details

Knowledge of Python programming for laboratory exercise is a prerequisite. Students are expected to recall the fundamental theory concepts relevant to the exercise to be performed in the upcoming laboratory.

The following 10 programming exercises and one design of lab problem statement will form the submission for laboratory coursework.

Sr. No.	Week No.#	List of Lab Exercises	Mapped CO
1	1.	Study of different python packages like numpy, pandas and matplotlib.	CO1
2	2.	Perform Exploratory Data Analysis on real world dataset using Pandas library i. Read different types of data files(csv, excel, text file etc.) ii. Obtain metadata of given dataset iii. Handling Missing Values in dataset iv. Handling Categorical data: Mapping ordinal features, Encoding class labels, Performing one-hot encoding on nominal features v. Variable Transformation and Variable Creation vi. Selecting meaningful features	CO1
3	3	Analyze given data using Data Visualization techniques using Matplotlib and identify trends, patterns and outliers with the help of different types of graphs.	CO1
4	4	Apply pre-processing on given dataset using scikit-learn package to transform raw features into representations suitable for machine learning model.	CO1
5	5	Implement Simple and Multiple Linear Regression on real world dataset and estimate the parameters of regression. Analyze the effect of varying learning rate and number of iterations i. Implement Gradient Descent algorithm using Numpy on toy dataset and observe the effect of varying learning rate and number of iterations. ii. Implement Simple and Multiple Linear Regression using scikit learn iii. Examine the effect of penalizing the parameters	CO2
6	6	Implement Logistic Regression on real world using sklearn, calculate class probability and evaluate the performance using confusion matrix.	CO2
7	7	Implement Decision tree classification algorithm and Ensemble learning algorithm (Random Forest, AdaBoost, Xgboost) i. Implement decision tree algorithm and identify and reduce the overfitting if it occurs. ii. Compare and analyse the output of Decision tree	CO3

		algorithm with ensemble learning algorithm	
8	8	Implement Support Vector Machine Classification algorithm a. Measure accuracy by varying kernels b. Tune the model using regularization and gamma parameter	CO3
9	9	Implement K means clustering and analyse the effect of varying the number of clusters	CO3
10	10	Implement the recommendation system by using nearest neighbour.	
Note: These are indicative lab exercises and may change time to time.			

5 Assessment Policy

5.1 Component wise Continuous Evaluation Internal Continuous Assessment (ICA) and Term End Examination (TEE)

Assessment Component	**ICA (100 Marks) (Marks scaled to 50)**				
	Lab Performance	**Lab Viva + Mini Project**	**Assignment-1**	**Class Test1 and Class Test 2**	**Assignment-2 Research Paper/ Literature Survey on Machine Learning**
Weightage	10%	10%	5%	20%	5%
Marks	20	20	10	20+20	10
Date/week	Weekly	Week 12 and Week 13	Week 10 and Week 11	Test-1 August, 20-23 August,2023 Test-2 October 10-15 October,2023	Regular Basis
CO Mapping	CO1, CO2, CO3, CO4	CO1, CO2, CO3,CO4	CO1, CO2, CO3,CO4	CO1, CO2, CO3,CO4	CO1, CO2, CO3, CO4

5.2 Assessment Policy for Internal Continuous Assessment (ICA)

Assessment of ICA comprises of the following components.

1. **Class test 1 and 2**

   a. Two class tests will be conducted as per the academic calendar.

b. It may be conducted online/ offline for 20 marks each

2. Lab performance and viva evaluation (20 marks)

a. Lab experiments (15 marks)
   i. Continuous assessment for laboratory experiments will be conducted.. There are 10 practical's, each carrying weightage of 10 marks. At the end of the course, average of total marks will be taken to obtain marks out of 10.
   ii. Discussion of your work with your peers is allowed. However, each student is expected to submit his/her original work. Submissions which are very similar will be marked zero. Assessment of the lab work will be carried out based on parameters like timely completion of lab work file, understanding of the experiment performed, originality in the work, involvement of the student, regularity, discipline etc. during the session. There is a 30% penalty on late submission.

b. Viva(5marks)
   i. Lab viva will be conducted based on the concepts learnt in the lab.

3. Mini Project (20 marks) –

- Student should form a group of 2-3 students and implement Mini Project
- Topic should cover EDA, regression, classification and clustering.

4. Assignment: **(10 marks)-**
Faculty can assign assignment based on theory, numerical or implementation.

5. Class Participation (10 marks)- The faculty will ask some questions in every class based on the content being taught. The question could be asked to a chosen student or a student group (which is formed at the beginning of the semester). One mark can be given to the correct answer. The idea is to encourage students to pay attention in class and actively participate. These marks will be added in ICA class participation component.

5.3 Assessment Policy for Term End Examination (TEE)

A written examination of 100 marks for 3 hours duration will be conducted for the course as per the academic calendar.

6. Lesson Plan

Session No.	Topics	Mapped CO	Reference
1	**Machine Learning Fundamentals** Terminology, Supervised and Unsupervised Learning with examples,	1	TB1

2	Underfitting / Overfitting, Bias-Variance Trade-off, Model Selection, Applications	1	TB1
3	**Exploratory data Analysis** Missing Value Treatment, Handling Categorical data: Mapping ordinal features, Encoding class labels, Performing one-hot encoding on nominal features	1	TB1
4	Outlier Detection and Treatment. Feature Engineering: Variable Transformation and Variable Creation, Selecting meaningful features.	1	TB2,Lab Manual + related internet articles
6	**Regression** Linear regression using Least Squares (analytical approach),	1	TB2,Lab Manual + related internet articles
7	Linear regression using Least Squares (analytical approach),		
8	Linear regression using Gradient Descent (iterative approach)(continued)	1	TB2,Lab Manual + related internet articles
9	Multiple linear regression,	1	TB2,Lab Manual + related internet articles
10	Polynomial regression	1	TB1
12	**Classification** Performance Evaluation, Confusion Matrix, Accuracy, Precision, Recall, F1-score,	2	TB1
14	ROC Curves, AUC, k-fold Cross-Validation. Logistic Regression, Naive Bayes Classifier	2	TB1
16	Support Vector Machines, Neural Networks: Perceptron	2	TB1
18	Multi-layer Perceptron, Training using Back-propagation, Applications	2	TB1
20	**Tree-Based Methods** Basics of Decision Trees, Regression Trees, Classification Trees	2	TB1
22	Trees v/s Linear Models, Advantages and Disadvantages of Trees,	2	TB1
24	Ensemble techniques- Bagging, Boosting, Random forest	2	TB1
26	**Unsupervised Learning** Challenges of Unsupervised Learning, Principal Component Analysis (PCA),	2	TB1
28	Partitioning Methods: K-Means clustering,	2	TB1

	Dimension Reduction Methods,		
30	Hierarchical Clustering, Introduction to Recommender systems (Non-personalized and Content-based)	2	TB1

1 Teaching-learning methodology

Faculty will make a group of 2-3 students for any group based activity such as class participation, project, presentation etc. Lecture and laboratory session will be conducted as follows-

1. **Lectures:**
   - Outline for preliminary study to be done for each unit will be provided prior to commencement of each unit.
   - Deeper concepts and applications will be explained through Presentation and Video Lectures.
   - Numerical problems based on concept will be solved during the session on *smart board* or *MS OneNote.*
   - Some practical applications will be simulated in class on *Jupyter notebook or Google Colab* for better understanding of the concepts.

2. **Laboratory:**
   - Lab manual consisting of theory and algorithm to support the lab experiment will be uploaded on student portal.
   - Regular lab assessment and grading will be done. Students will be marked based on parameters like completion of lab assignment, originality, logic developed, interaction during the lab, submission, punctuality and discipline

8 Active learning techniques

Active learning is a method of learning in which students are actively or experientially involved in the learning process. Following active learning techniques will be adopted for the course.

Active learning is a method of learning in which students are actively learning process. Following active learning techniques will be adopted for the course.

1. Padlet Interaction/Google Form: Students will be asked to write their views about the topic learned in the class on padlet. Students will also be encouraged to comments of peers' inputs and gradethem.

2.Think pair and Share:

T (Think): Teacher asks a specific question about the topic. Students "think" about what they know or have learned, and come up with their own individual answer to the question. [Takes 1-3 Minutes].

P (Pair): Teacher asks another question, related to the previous one, that is suitable to deepen the students' understanding of the topic. Each student is paired with another

student. They share their thinking with each other and proceed with the task. [Takes 5-10 Minutes].

S (Share): Students share their thinking (or solution) with the entire class. Teacher moderates the discussion and highlights important points. [Takes 10-20 minutes].

3.Critical Thinking development:

Critical discussion by doing so, it aims at lowering the barrier for researchers and educators to understand technical concepts and consider critical issues around machine learning.

9 Course Material

Following course material is uploaded on the student portal: (give student portal link)

- Course Policy
- Lecture Notes
- Lecture Presentations
- Books / Reference Books
- Assignments and case studies
- Lab Manuals, Dataset links
- List of Program Outcomes

10 Course Outcome Attainment

Following means will be used to assess attainment of course learning outcomes.

- Use of formal evaluation components of continuous evaluation, assignments, laboratory work, semester end examination
- Informal feedback during course conduction

11. Academic Integrity Statement

Students are expected to carry out assigned work under Internal Continuous Assessment (ICA) independently. Copying in any form is not acceptable and will invite strict disciplinary action. Evaluation of corresponding component will be affected proportionately in such cases. Plagiarism detection software will be used to check plagiarism wherever applicable. Academic integrity is expected from students in all components of course assessment.